(2)

AD-A198 588

REPORT DOCUMENTATION PAGE

Unclassified	1b. RESTRICTIVE MARKINGS DTIC FILE COPY
2a. SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NMRI 87-82	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Medical Research 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Naval Medical Command
6c. ADDRESS (City, State, and ZIP Code) Bethesda, Maryland 20814-5055	7b. ADDRESS (City, State, and ZIP Code) Department of the Navy Washington, D.C. 20372-5120
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Naval Medical Research and Development Command 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Bethesda, Maryland 20814-5055	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61153N; PROJECT NO. MR041.20; TASK NO. MR0412003 0002; WORK UNIT ACCESSION NO. DN377032

11. TITLE (Include Security Classification)

A PREDICTIVE MODEL OF ENZYMATIC CLEAVAGE OF NUCLEIC ACIDS

12. PERSONAL AUTHOR(S)

CHRISTOPHER J. HOUGH

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Report No. 6	FROM _____ TO _____	1987	19

16. SUPPLEMENTARY NOTATION

Reprinted from: Biopolymers, Vol. 26, 1789-1807 1987

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	DEOXYRIBONUCLEASES	MODELS, THEORETICAL
			DNA	RIBONUCLEASES
			MICROCOCCAL NUCLEASE	RNA

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Phyllis Blum, Information Services Division	22b. TELEPHONE (Include Area Code) 202-295-2188 22c. OFFICE SYMBOL ISD/ADMIN/NMRI

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

1729

A Predictive Model of Enzymatic Cleavage of Nucleic Acids

CHRISTOPHER J. HOUGH, *Transplantation Research Program Center, Biochemistry Branch, Naval Medical Research Institute, Naval Medical Command National Capitol Region Bethesda, Maryland 20814*

Synopsis

A new static model of enzymatic cleavage of polymeric substrates such as nucleic acids has been derived. The model is compared to that elaborated by Tanford and to experimental data. In predicting the fragment distribution for a restriction enzyme digestion of a circular substrate, the model is superior to that of Tanford.

INTRODUCTION

The use of endonucleases for the manipulation of small discrete nucleic acid sequences and the probing of the structure of large complex chromosomes has become the basis of experimental design in molecular biology. Despite widespread use of these tools, little is known about their mechanisms of action and kinetics. For example, the effect of base sequence surrounding a restriction enzyme site on the rate of cleavage of the site is not well known. This is in part due to the complexity inherent in digestions of even simpler substrates. In chromatin studies, where endonuclease digestions are an integral part of obtaining DNA sequences of functional interest and where the spacial relationship of associated molecules to a particular nucleic acid sequence is of concern, the number of possible digestion products is enormous.

Hence, there is a need for a model of enzymatic cleavage of nucleic acid substrates to determine what possible digestion products might be expected, to study the effects of site susceptibilities on the rate of appearance of these products, or to determine the expected distribution of fragments containing both nucleic acid sequence and associated chromatin proteins of interest. These expectations can in turn be compared to experimental data to derive meaningful conclusions.

One such model, that given by Flory[1] and elaborated for use in polymer degradation reactions by Tanford,[2] has served as the basis for most models used to date. In this model the creation or cleavage of each bond is considered as occurring independently of other bonds or the polymer chains to which it belongs. Polymerization or degradation is expressed in terms of degree, or fraction, of bonds created or broken, a quantity that can easily be measured by following the appearance or disappearance of small monomeric reactants or products. This was appropriate for the chemical polymerization of small organic monomers. However, for the enzymatic degradation of DNA or RNA, this perspective is unsatisfying for several reasons.

Biopolymers, Vol. 26, 1789–1807 (1987)

CCC 0006-3525/87/101789-19$04.00

First, the extent of reaction is most often followed by observing the fragments resulting from the digestion. Second, for enzymatic digestions of nucleic acids that have been studied in detail,[3] the kinetics of bond cleavage have manifested a dependence on the polymer chain to which the bond belongs. Mechanisms of facilitated diffusion or chain processing have been proposed to explain these data.[4]

A model is proposed here that approaches the degradation of a polymer chain from the perspective of resulting fragments. In so doing, the *possibilities* of bond cleavage within one polymer chain are depicted as representing the digestion as a whole. Although this model represents simply another perspective of polymer degradation, it is a view that may be more amenable for the study of enzymatic digestions of DNA or RNA.

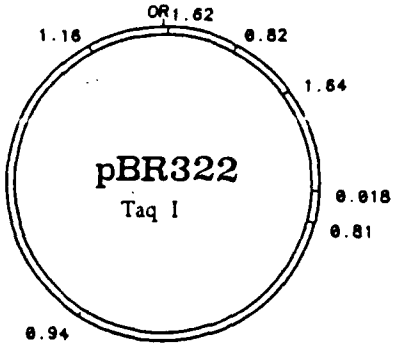
MATERIALS AND METHODS

Restriction Enzyme Digestion

pBR322 (Boehringer Mannheim) was digested for 1-10, 15, 30, 45, and 90 min with 1 and 10 units/μg DNA of Taq I (BRL) at 65°C. The digestion was stopped by the addition of 10 mM EDTA, NaCl to 0.25M, and 2.5 vol ethanol. The DNA was precipitated from the ethanol, dried, and run on a 1.75% agarose gel in the presence of 2 μg/mL ethidium bromide. The gel was photographed and the negative was scanned by laser densitometry (LKB model 2202 ultrascan). Assignment of peaks was made by comparing the relative migration on each to standards (Hind III digest of λ-DNA). The time at which each restriction fragment reached 2/7ths of the total by number, $t_{2/7}$, was interpolated from this data, in every case within a reasonably linear region of the time course of appearance of each fragment. These times were, for each fragment in decending order by size, 7.72, 11.05, 280.91, 4.46, 6.31, 6.26, and 569.52 min. The fragments of 315 and 312 base pairs (bp), although not resolved on a 1.75% agarose gel, could be distinguished by a shift in band mobility toward higher molecular weight of the comigrating fragments, the disappearance of the 928bp partial fragment containing the 312-bp unit fragment, and the disappearance of the 683-bp partial fragment containing the 315-bp unit fragment (as seen by a shift toward lower molecular weight of the comigrating 613-bp partial fragment). These observations indicated that the production of the 315-bp fragment lagged slightly behind that of the 312-bp fragment. The $t_{2/7}$ given for these two fragments represent an estimate of this difference. Partial fragments of 1234 and 1448 bp were also not resolved from those of 1307 and 1444 bp, respectively. The rates of appearance of the 1307- and 1444-bp fragments were in all probability overestimated. The reciprocal of these times were made proportional to the product of the two restriction sites at the two ends of each fragment. From the seven equations of seven unknowns, the site susceptibility of each site was then calculated. These susceptibilities were also constrained such that their sum was equal to 7. The calculated site susceptibility for each cleavage site is indicated in Fig. 1.

Micrococcal Nuclease Digestion

Human bone marrow cells were obtained fresh from the Bethesda Naval Hospital tissue bank. Human bone marrow nuclei were prepared by the



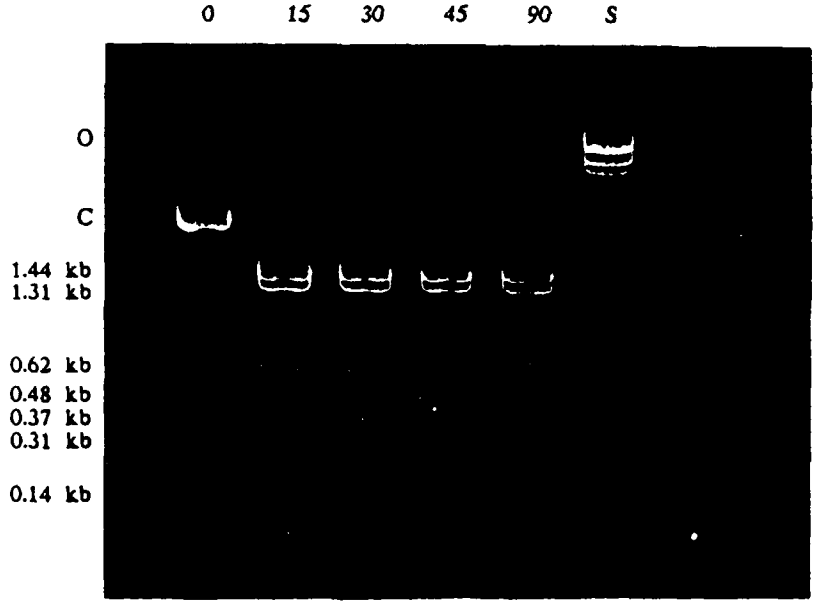


Fig. 1. Taq I digest of pBR322. The Taq I restriction map of pBR322 is given with the experimentally obtained (see the materials and methods section) site susceptibilities indicated for each site. The fragments obtained from a complete digest are, proceeding clockwise from the origin, 315, 313, 475, 141, 1307, 1444, and 368 base pairs in size. Below the restriction map is shown a Taq I digest of pBR322 (10 unit/μg) at 65°C for 0, 15, 30, 45, and 90 min electrophoresed on a 1.75% agarose gel containing 2 μg/mL ethidium bromide. The Hind III digested λ-DNA standards are indicated by the S. The size assignments of the major bands, O representing the open form, C representing the supercoiled circular form, and the fragment sizes in kilobases, were determined by comparison to the λ-DNA standards. The 141-bp fragment of pBR322 and the 125-bp fragment of λ are scarcely visible near the bottom of the photograph.

method of Wallace et al.[5] with few modifications. Triton X-100 to 0.05% was added to the homogenization buffer immediately prior to homogenization. The nuclei were then treated for 3 min with 50 units/mg DNA of Micrococcal nuclease at room temperature, as described by Davis et al.[6] The digestion was stopped by the addition of 2 mM EGTA and chilling. Total digestion products were isolated by precipitation from 6M guanadinium hydrochloride with 2 vol of ethanol at −20°C overnight, and further purified as previously described.[7] A fractionation of the digestion products was achieved by centrifuging the treated nuclei at 10,000 × g, resuspending in 1 mM EDTA, 1 mM EGTA for 1 h at 0°C, and recentrifuging. The DNA remaining in the pellet was isolated as above. Purified digestion products were electrophoresed on a 1.5% agarose gel. The gel stained with 2 μg/mL ethidium bromide was photographed, and the negative was scanned by densitometry. The degree of digestion was estimated by taking the square root of the weight fraction of the mononucleosomal peak.

RESULTS

A model of enzymatic digestion of nucleic acids was approached by asking the question, How many ways is it possible to cleave a substrate to produce a given fragment length? Because circular substrates are somewhat different from linear substrates in this regard, the two kinds of substrates are presented separately. The problem is initially approached from the standpoint that all cleavages are equally susceptible. Later, different susceptibilities of sites are considered within the framework established. The model is then compared to that of Tanford.

Sites of Equal Susceptibility

Linear Substrates. Let the substrate be a linear polymer of N units. These units need not be of identical length or size, but the units must be joined by identical bonds that happen to be cleavable by an enzyme of interest. The length in monomer units of a fragment of the substrate resulting from the action of the enzyme on the substrate is L, again possibly of varying sizes but nevertheless composed of L units. There are $N - 1$ sites of cleavage in a linear substrate. To produce a fragment of length L, cleavage must occur at either end of, but not within, the fragment. Production of any fragment within the substrate requires two cleavages, unless the fragment resides at either end of a linear substrate where only one cleavage is required. Therefore, to produce a fragment of length L from a linear substrate, only $N - L - 2$ sites remain for additional cleavage in the case of interior fragments and $N - L - 1$ sites in the case of terminal fragments.

If the substrate is cleaved at x sites with equal probability, then what is the number of combinations of cleavages possible that result in a given fragment of length L? For fragments in the interior of a linear polymer, this is simply the number of ways $x - 2$ cleavages—the number of cleavages remaining after cleaving a given internal fragment—can be arranged among the available $N - L - 2$ sites lying outside of the fragment, or

$$C(N - L - 2, x - 2)$$

where $C(n, p) = n!/p!(n-p)!$. Similarly, for fragments at the ends, the combinations of cleavages resulting in a given fragment of length L is

$$C(N - L - 1, x - 1)$$

Since the number of different products of length L is $N - L + 1$ from the digestion of a linear substrate, the total number of combinations resulting in all L length fragments from x cleavages of a linear substrate is

$$2 \cdot C(N - L - 1, x - 1) + (N - L - 1) \cdot C(N - L - 2, x - 2)$$

$$= (x + 1) \cdot C(N - L - 1, x - 1)$$

Note that, since $x + 1$ represents the number of fragments of *any* length created by x cleavages, the total number of combinations resulting in L length fragments from x cleavages is the same as the product of the number of combinations resulting in a *given* fragment at a terminus of the substrate and the number of fragments of any length created by x cleavages on a single substrate molecule. In a collection of substrate molecules representing every possible combination of x cleavages, then, the above expression represents the number of fragments of length L produced.

A point of definition must be clarified. The combinatorial equation,

$$C(n, p) = \frac{n!}{p!(n-p)!}$$

is defined for $n = p = 0$, but it is not defined for $p > n$ and n or $p < 0$. In the model described above, however, it is possible to imagine a substrate cleaved by an enzyme to the point that a resulting fragment of length $L > 1$ is not possible, or conversely, that a fragment of length $L < N$ is not possible from one or no cleavage. Such cases are described by $x - 1 > N - L - 1$ and $N - L - 1$ or $x - 1 < 0$. Thus, it is imperative that, for this model, $C(n, p)$ be defined as 0 when $p > n$ and n or $p < 0$.

The total number of combinations of x cleavages within a linear substrate is $C(N - 1, x)$, and the number of fragments obtained from all such combinations is $(x + 1)C(N - 1, x)$. Thus, the number fraction F_{no} of all fragments of length L resulting from every combination of x cleavages of a linear substrate is

$$F_{no} = \frac{C(N - L - 1, x - 1)}{C(N - 1, x)} \tag{1}$$

The right-hand expression in Eq. (1) is, in essence, the probability of producing L length fragments from x cleavages. Provided every cleavage is equally likely, this probability should hold, whether or not every combination of x cleavages has been explored. The corresponding weight fraction is the total number of fragments of length L times the fraction by weight of the total

substrate represented by such a fragment

$$F_{wt} = \frac{L}{N} \cdot (x+1) \cdot \frac{C(N-L-1, x-1)}{C(N-1, x)} \tag{2}$$

where the length of all fragments in polymer units is the same.

Where fragments of length L differ in molecular weight (contain different numbers of polymer units), the number fraction of any given fragment is represented by

$$F_{no} = \frac{1}{(x+1)} \cdot \frac{C(N-L-2, x-2)}{C(N-1, x)} \tag{1a}$$

for internal fragments, where L retains the definition given above, and

$$F_{no} = \frac{1}{(x+1)} \cdot \frac{C(N-L-1, x-1)}{C(N-1, x)} \tag{1b}$$

for terminal fragments. If l is the length in polymer units of a given fragment of length L, and n is the length of the substrate in polymer units, then the weight fraction of any given fragment is

$$F_{wt} = \frac{l}{n} \cdot \frac{C(N-L-2, x-2)}{C(N-1, x)} \tag{2a}$$

for internal fragments, and

$$F_{wt} = \frac{l}{n} \cdot \frac{C(N-L-1, x-1)}{C(N-1, x)} \tag{2b}$$

for terminal fragments.

The degree of digestion, $x/N-1$, can be determined from the weight fraction using a rearrangement of Eq. (2) (assuming identical L length fragments):

$$\frac{x}{N-1} = \frac{N}{x+1} \cdot \frac{F_{wt}}{L} \cdot \prod_{i=1}^{x-1} \frac{N-1-i}{N-L-i}, \qquad x > 1 \tag{3}$$

For large x and N, Eq. (3) simplifies, when $L = 1$, to

$$\frac{x}{N-1} = \sqrt{F_{wt}}, \qquad L = 1 \tag{4}$$

Similar treatments of Eqs. (1a) and (1b) yield

$$\frac{x}{N-1} = \frac{N-L-1}{x-1} \cdot F_{wt} \cdot \frac{n}{l} \cdot \prod_{i=1}^{x-1} \frac{N-1-i}{N-L-i}, \qquad x > 1 \tag{3a}$$

for a given internal fragment, and

$$\frac{x}{N-1} = F_{wt} \cdot \frac{n}{l} \cdot \prod_{i=1}^{x-1} \frac{N-1-i}{N-L-i}, \qquad x > 1 \tag{3b}$$

for a given terminal fragment. Analogous to Eq. (4), the degree of digestion for large x and N when $L = 1$ can be expressed as

$$\frac{x}{N-1} = \sqrt{F_{wt} \cdot n/l}, \qquad x > 1 \tag{4a}$$

for internal fragments, and

$$\frac{x}{N-1} = F_{wt} \cdot n/l, \qquad x > 1 \tag{4b}$$

for terminal fragments.

Circular Substrates

In circular substrates there are N sites of cleavage, and all fragments, except the open circle, must be generated by two cleavages. Thus, the number of combinations resulting in a given fragment of length $L < N$ from x cleavages is

$$C(N-L-1, x-2)$$

where $N - L - 1$ is the number of sites available for cleavage apart from the two required to cleave the fragment, and $x - 2$ is the number of cleavages remaining after cleavage of the fragment. Similarly, with one cleavage, $C(L - N, x - 1)$ describes the fact that there is only one combination giving the open circle and that only a fragment of $L = N$ is possible. Since there are N possible fragments of length L in a circular substrate, the number of fragments of length L resulting from every combination of $x > 1$ cleavages is

$$N \cdot C(N-L-1, x-2)$$

For a single cleavage, only one "fragment" of length N can be produced, but there are N such fragments possible (one for each permutation of the sequence). Thus the number of different fragments of length N possible from a single cleavage is

$$N \cdot C(L-N, x-1) = N$$

The total number of combinations of cleavages within a circular substrate is $C(N, x)$. The number of fragments of *any* length resulting from any one combination of x cleavages equals x. Therefore the number fraction of fragments of length L resulting from $x > 1$ cleavages of a circular substrate is expressed by Eq. (5):

$$F_{no} = \frac{N}{x} \cdot \frac{C(N-L-1, x-2)}{C(N, x)} \tag{5}$$

For $x = 1$ the equivalent to the right-hand side of Eq. (5) is

$$\frac{N}{x} \cdot \frac{C(L-N, x-1)}{C(N, x)}$$

which is nonzero only for $x = 1$ and $L = N$, and then it is equal to 1. The corresponding weight fraction is

$$F_{wt} = L \cdot \frac{C(N-L-1, x-2)}{C(N, x)}, \qquad x > 1 \tag{6}$$

Here again, the equivalent expression for a single cleavage, $L \cdot C(L-N, x-1)/C(N, x)$, reduces to 1.

For any given fragment within the circular substrate, the number fraction is

$$F_{no} = \frac{1}{x} \cdot \frac{C(N-L-1, x-2)}{C(N, x)} \tag{5a}$$

for $x > 1$, and

$$F_{no} = \frac{1}{x} \cdot \frac{C(L-N, x-1)}{C(N, x)} = \frac{1}{N} \tag{5b}$$

for $x = 1$. The corresponding weight fraction for any given fragment of length L cleavage units but length l polymer units is

$$F_{wt} = \frac{l}{n} \cdot \frac{C(N-L-1, x-2)}{C(N, x)} \tag{6a}$$

for $x > 1$, and

$$F_{wt} = \frac{n}{n} \cdot \frac{C(L-N, x-1)}{C(N, x)} = \frac{1}{N} \tag{6b}$$

for $x = 1$, where n is the length of the substrate in polymer units.

Rearranging in the same way as in Eq. (4), the degree of digestion x/N is given by Eq. (7):

$$\frac{x}{N} = \frac{N-1}{x-1} \cdot \frac{F_{wt}}{L} \cdot \prod_{i=1}^{x-2} \frac{N-1-i}{N-L-i}, \qquad x > 2 \tag{7}$$

For large x, large N, and $L = 1$, Eq. (7) reduces to

$$\frac{x}{N} = \sqrt{F_{wt}}, \qquad L = 1 \tag{8}$$

Rearranging Eq. (6a), the degree of digestion can be expressed in terms of the weight fraction of any given fragment smaller than N by

$$\frac{x}{N} = \frac{N-1}{x-1} \cdot F_{wt} \cdot \frac{n}{l} \cdot \prod_{i=1}^{x-2} \frac{N-1-i}{N-L-i}, \qquad x > 2 \tag{7a}$$

This simplifies for large x and N when $L = 1$ to

$$\frac{x}{N} \approx \sqrt{F_{wt} \cdot n/l}, \qquad L = 1 \tag{8a}$$

Sites of Unequal Susceptibility

In terms of the model outlined above, a cleavage site having, for example, twice the cleavage susceptibility of another site means that the more susceptible site will be cleaved by a single cleavage in two substrate molecules for every one substrate molecule cleaved at the less susceptible site. Mathematically, this is expressed as a weighting of those combinations of cleavages containing sites of higher or lower susceptibility than the bulk of the sites of average susceptibility.

For any given fragment of length L, three kinds of sites may be distinguished. The abundance of this fragment among all those derived from the digestion of a substrate is affected differently by the susceptibilities of each. The sites most directly affecting the occurrence of a particular fragment are those at either boundary of the fragment. Those having an indirect effect on the occurrence of the fragment are those sites lying outside the boundaries of the fragment. This is due to the fact that every site within the substrate of susceptibility greater than 1 increases the number of substrate molecules necessary to describe every combination of cleavages. Those sites that lie within the fragment have no effect on the occurrence of that fragment in a digestion. Since a fragment requires one or two cleavages for its occurrence, the frequency of occurrence of the fragment is weighted by the product of the susceptibilities of those cleavages. If r is the number of cleavages required to produce the fragment and h_i is the susceptibility of the ith site, then the frequency of occurrence of the L length fragment spanning sites i and $i + L$ is weighted by the product

$$A_r = \prod_{j=1}^{r} [h_i],$$

where i represents the boundary sites of the fragment.

This leaves $x - r = k$ cleavages available for sites outside the boundaries of the fragment. These occur in every possible combination over all the sites lying outside the boundaries of the fragment. If there are $b = N - L - r$ such sites, then there are $C(b, k) = C(N - L - r, x - r)$ possible combinations of cleavages. Each combination is weighted by the product of all k site susceptibilities. The sum of these products, then, represents the effect of sites lying outside the boundaries of the fragment. Let E_k represent this sum, such that

$$E_k = \begin{cases} \sum_{n=1}^{C(b,k)} \left[\prod_{j=1}^{k} (h_i)_j \right]_n, & k = x - r > 0 \\ 1, & k = x - r \le 0 \end{cases}$$

where i represents each of the sites lying outside the boundaries of the fragment in all combinations taken k at a time. Having weighted each combination of the cleavages that is capable of producing a particular fragment, the expression representing what fragment sizes are possible for the linear substrate becomes

$$C(N - L - r - k, x - r - k) = C(N - L - x, 0)$$

For circular substrates the equivalent expression is

$$C(N - L - x + 1, 0)$$

These can assume the values of 1 or 0. Thus, the number of combinations resulting in a given fragment of length L from x cleavages is

$$A_r \cdot E_k \cdot C(N - L - x, 0)$$

for linear substrates or

$$A_r \cdot E_k \cdot C(N - L - x + 1, 0)$$

for circular substrates. That these more general expressions for the number of combinations of x cleavages resulting in a fragment of length L are consistent with those derived for equal site susceptibilities can be seen by making all h_is equal to 1.

To determine the number of all such L length fragments, this expression must be evaluated separately for each of the $N - L + 1$ (linear substrate) or N (circular) possible fragments and summed. This sum can then be evaluated for each possible L and again summed to obtain the number of fragments of any length resulting from x cleavages. Though inconvenient to express mathematically, this series of operations can be done by computer. Programs for this purpose have been written for both linear and circular substrates. Since the frequency of occurrence of each individual fragment must be calculated, the program has been written so that any portion or portions of the substrate can be earmarked and followed through the course of digestion. This permits, for example, the modeling of the fate of a gene or the distribution of fragments containing two loci of interest within the substrate during enzymatic digestion.

Underlying Assumptions

In essence, the model above describes the number of combinations of random cleavages of a substrate molecule that results in a particular fragment or fragment length (in unit fragments). When applying this model to real digestions of DNA or RNA substrates, a number of underlying assumptions must be made. It is assumed that the cleavage of sites within each substrate molecule occurs randomly, that each combination of cleavages is represented by the cleavages taking place on a single substrate molecule, and that each substrate molecule is initially identical. These assumptions are frequently justified in digestions of DNA and RNA substrates.

Since each substrate molecule is considered a representative of one combination of cleavages among those possible, it is also assumed that each molecule is cleaved to the same degree at any point in time. This assumption is clearly difficult to justify in reality. The model may be extended, however, to consider the products of a digestion as a distribution of varying degrees of cleavage about a mean, each degree of cleavage represented by a collection of molecules present in all combinations of cleavages. The continuous nature of a digestion in reality is then represented as a statistical distribution of many discontinuous events, the formalism for which can be drawn from classical statistical mechanics.

Comparison to Tanford's Model

The model of polymer cleavage developed here differs from that of Tanford in subtle but important ways. In following the development of Flory,[1] Tanford treats polymer chain degradation as the reverse of polymerization. As such, the bonds joining monomeric units are considered a collection, irrespective of position or substrate molecule. Chain cleavage is then represented as a random breaking of the bonds in the collection. Since polymer cleavage can often be followed by a change in pH or disappearance of reactant, this treatment is developed in terms of bonds broken. The degree of depolymerization is represented as the fraction q of all bonds present that are broken. There is no difference in treatment in the Tanford model for linear or circular substrates. In contrast, the present model considers each substrate molecule as a subset of the total digestion and the collection of substrate molecules as the universe representing every possible combination of random cleavages.

For equal site susceptibilities, the Tanford treatment predicts that the number fraction of all fragments of length L created by qN_0 cleavages, where N_0 is the number of bonds present, is

$$q(1-q)^{L-1}$$

When $N_0 = N - 1$ and $q = x/(N-1)$ (assuming a linear substrate), this expression becomes

$$\frac{x}{(N-1)}\left[\frac{N-x-1}{N-1}\right]^{L-1}$$

very similar to the corresponding expression in the present model,

$$\frac{C(N-L-1,x-1)}{C(N-1,x)} = \frac{x}{(N-1)}\prod_{i=1}^{L-1}\left[\frac{N-x-i}{N-1-i}\right]$$

Both predict that, at $L = 1$, the number fraction of fragments is $x/(N-1)$. For large N their difference is negligible. Note, however, that the number fraction of uncleaved substrate remains finite through the entire digestion in the Tanford model but not in the present model. This is a direct consequence of the difference between the two models noted above.

In models following the Tanford treatment that include differing site susceptibilities,[8] the treatment of these susceptibilities also differs from the present model in important ways. In the Tanford model, the fraction of bonds broken during digestion, q, can be considered the probability that a cleavage

will occur at any given site. As such, the probability of occurrence of any given fragment is the product of the likelihood of cleavage at each boundary site and the likelihood that cleavage will not occur, $(1 - q)$, at each of the sites within the fragment, hence,

$$F_{wt} = q^2 \cdot (1 - q)^{L-1} \tag{9}$$

Cleavages of unusual susceptibility have been considered, plausibly, as cleavages of greater or less probability. This probability has been expressed as Sq, where S represents the factor of greater or less susceptibility. For this model to be valid, however, q must represent the fraction of bonds broken, such that qN_0 equals the number of broken bonds or cleavages. If one bond is represented as having Sq probability of cleavage while the remainder have q probability of cleavage, the number of possible cleavage sites is increased by $S - 1$, and q no longer represents the fraction of cleaved sites. In effect, a single site is represented by S sites, making the site a larger target for cleavage, whence the term *target analysis* of sites has been used.[8] In the present model, the concept of site susceptibility has been represented as a weighting of those combinations of cleavages that result in a given fragment. The effect of doing so is to increase the number of substrate molecules necessary to represent every combination possible while retaining the intrinsic number of sites present.

The meaning of q may be preserved by making the remaining "ordinary" sites slightly less susceptible than 1, such that

$$\sum_{i=1}^{N_0} S_i = N_0$$

For large N_0 and few unusual cleavage sites, the change in susceptibility of the ordinary cleavage sites can be negligible. The success of the application of the above treatment for digestions of DNA with DNase I may in part be due to the fact that the DNase I hypersensitive sites are indeed multiple sites created at an "unusual" locus of cleavage.[9]

COMPARISON TO EXPERIMENTAL DATA

To test the model against experimental data, three separate applications have been chosen. One is a restriction digest of a circular substrate (pBR322) in which site susceptibility plays an important role. The other two are Micrococcal nuclease digestions of human bone marrow chromatin. Both the latter are modeled assuming equal site susceptibilities for all sites, since only the distribution of the bulk of the DNA by ethidium bromide staining was examined. In each case, model predictions were also made using the Tanford model for comparison between models. Where required, the predicted weight fractions were calculated using an equation consistent with the Tanford model derived to include unequal site susceptibilities. This derivation states that the weight fraction of each fragment of real length l but of polymerization L is described by

$$F_{wt} = l \cdot S_a \cdot S_b \cdot q^2 \cdot \prod_{i=1}^{L-1} (1 - S_i q) \tag{10}$$

where sites a and b are fragment boundaries and sites i are those in the interior of the fragment.

Taq I Digest of pBR322

There are 7 Taq I sites in pBR322 mapped to 24, 339, 651, 1126, 1267, 2574, and 4018 nucleotides from the Eco R1 site making the fragments 315, 312, 475, 141, 1307, 1444, and 368 nucleotides in length, respectively (see Fig. 1). Initial digestions of pBR322 with Taq I indicated that complete digestion of the substrate was difficult to achieve, even with 10 units of enzyme per μg DNA at 65°C for a period of 90 min (Fig. 1). In particular, the appearance of the 475 and 141 nucleotide fragments lagged behind that of the others. This suggested that the site at nucleotide 1126 is less susceptible to cleavage than the other sites. To determine the relative susceptibility of this site, a series of digestions were performed to measure the initial appearance of each unit fragment. If every site were equally susceptible to cleavage, the number fraction of each unit fragment at the point in time at which two cleavages, on average, had occurred per polymer chain would reflect the product of each fragment's boundary site susceptibilities. For unequal susceptibilities, this is not the case. Provided the digestion proceeds at a reasonably linear rate, however, the time required for each fragment to reach 2/7ths of a complete digestion, its $t_{2/7}$, would be inversely proportional to the product of the boundary site susceptibilities of that unit fragment. Thus the interpolated times at which each restriction fragment reached 2/7ths of a complete digestion were used to determine the relative susceptibilities of its boundary sites. Using the site susceptibilities obtained by this method, predictions were calculated by both models that most closely represented a degree of digestion obtained experimentally (Fig. 2). Table I compares quantitatively the experimental data shown in Fig. 2 and the calculated models. The differences in predictions between the two models is considerable as expected for small N. In particular, the Tanford model predicted that a substantial fraction of the substrate remained uncleaved, whereas in both the experimental data and the predictions of the present model, none of the intact substrate remained. The experimental data and predictions made by the present model compared favorably.

Micrococcal Nuclease Digestion of Chromatin

Nuclei were prepared from human bone marrow and treated with Micrococcal nuclease for various times. The DNA released from these nuclei was then purified, run on agarose gels, stained with ethidium bromide, and photographed. A densitometeric scan of the digest is shown in Fig. 3. The degree of digestion was determined by taking the square root of the weight fraction of mononucleosomal DNA band determined by densitometry. Model digestions of a linear substrate of 100 nucleosomes were calculated using Eqs. (2) and (9) to approximate the digest (Fig. 3). Good agreement between experimental and model results was found in both cases. As can be seen in this example, the differences between the two models is small when N is large. Deviation of the predictions from the experimental results in the high molecular weight range may be explained by the small substrate size used for the calculated predic-

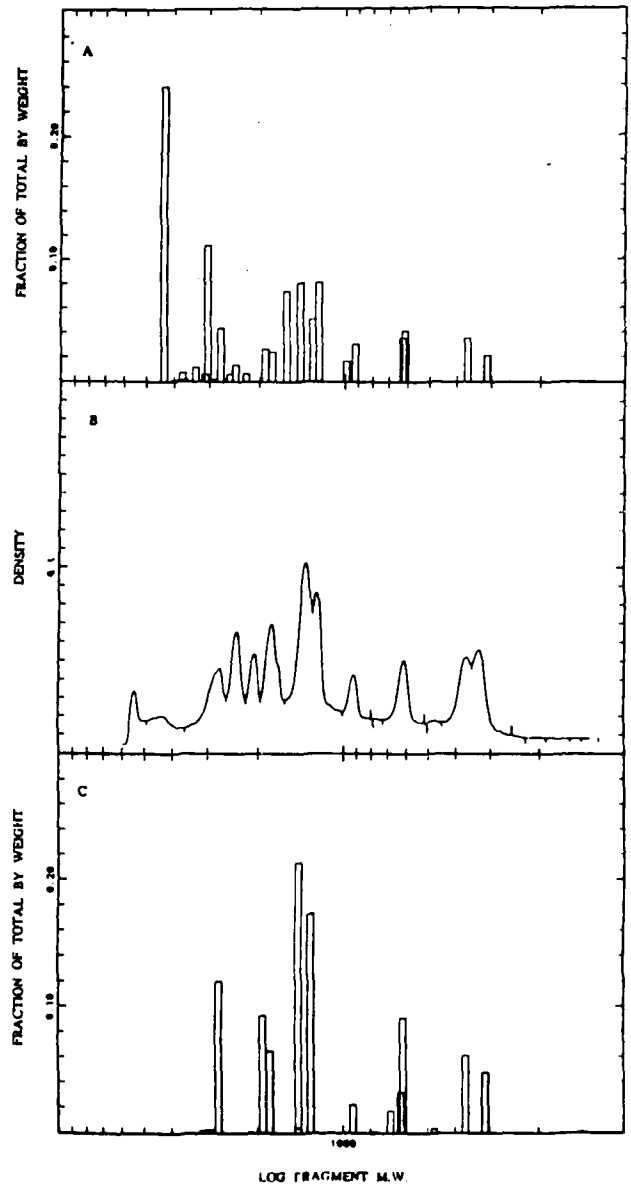


Fig. 2. Model predictions of a Taq I digest of pBR322. Panel A: predicted fragment distribution approximating the digest shown in panel B calculated using the site susceptibilities determined as described in the materials and methods section and Eq. (10) from the Tanford model. Panel B: A Taq I digest of pBR322. The plasmid pBR322 was digested with Taq I for 15 min at 65°C, electrophoresed, and scanned by densitometer as described in Fig. 1. The DNA Panel C: the predicted fragment distribution calculated as above using the present model. In all three panels the ordinate is inverted and scaled to the log of fragment size in base pairs to simulate patterns observed in agarose gel electrophoresis. The densitometric scan has been scaled to correspond, as nearly as possible, to the other two plots. A quantitative comparison of these data is given in Table I.

TABLE I
Quantitative Comparison of Model Predictions to Experimental Data

Predicted by		Tanford	Present	Experimental	
Rf[a]	Size	Fraction[c]	Fraction[c]	Rf[b]	Fraction
4.91	141	0.0001	0.0008	4.94	undetected
4.10	312	0.0215	0.0474	4.21	0.09316
4.09	315	0.0215	0.0478		
3.93	368	0.0357	0.0613	4.08	0.07710
3.67	475	0.0007	0.0033	3.67	0.00036
3.39	616	0.0416	0.0910	3.39	0.06923
3.36	627	0.0359	0.0318		
3.25	683		0.0173		
3.07	787		0.0013		undetected
2.86	928	0.0308	0.0222	2.86	0.04715
2.77	995	0.0169	0.0005		
2.64	1102		0.0006		
2.48	1243	0.0810			
2.42	1307	0.0511	0.1734	2.46	0.11047
2.29	1444	0.0808	0.2136	2.36	0.17466
2.28	1448	0.0005	0.0042		
1.99	1812	0.0247	0.0648	1.99	0.11387
1.92	1923	0.0272	0.0931		
1.77	2127		0.0008	1.82	0.06465
1.75	2235	0.0067			
1.68	2439	0.0140			
1.65	2550	0.0064		1.63	0.09330
1.59	2751	0.0439	0.1201		
1.56	2892	0.0003	0.0027		
1.55	2918	0.0022			
1.52	3055	0.1119			
1.51	3119	0.0064	0.0021		
1.48	3260	0.0000			
1.46	3367	0.0119		1.46	0.09356
1.40	3679	0.0021			
1.39	3735	0.0005			
1.38	3746	0.0077			
1.36	3887	0.0001			
1.34	3994	0.0015			
1.33	4047	0.0000			
1.28	4362	0.2404			

[a] Position interpolated from standards.
[b] Position and fraction of band determined by densitometry.
[c] Fraction of total sample by weight.

tions and by possible incomplete cleavage in some subset of the bone marrow nuclei. This latter possibility is explored below.

Complex Digestions

The use of the model to predict digestions containing subsets of varying degrees of digestion was explored by attempting to obtain a better prediction of the experimental results shown in Fig. 2B. In addition, the use of the model for predicting fragment distributions for cases in which a fractionation of digestion products has occurred by one means or another was tested. Using a

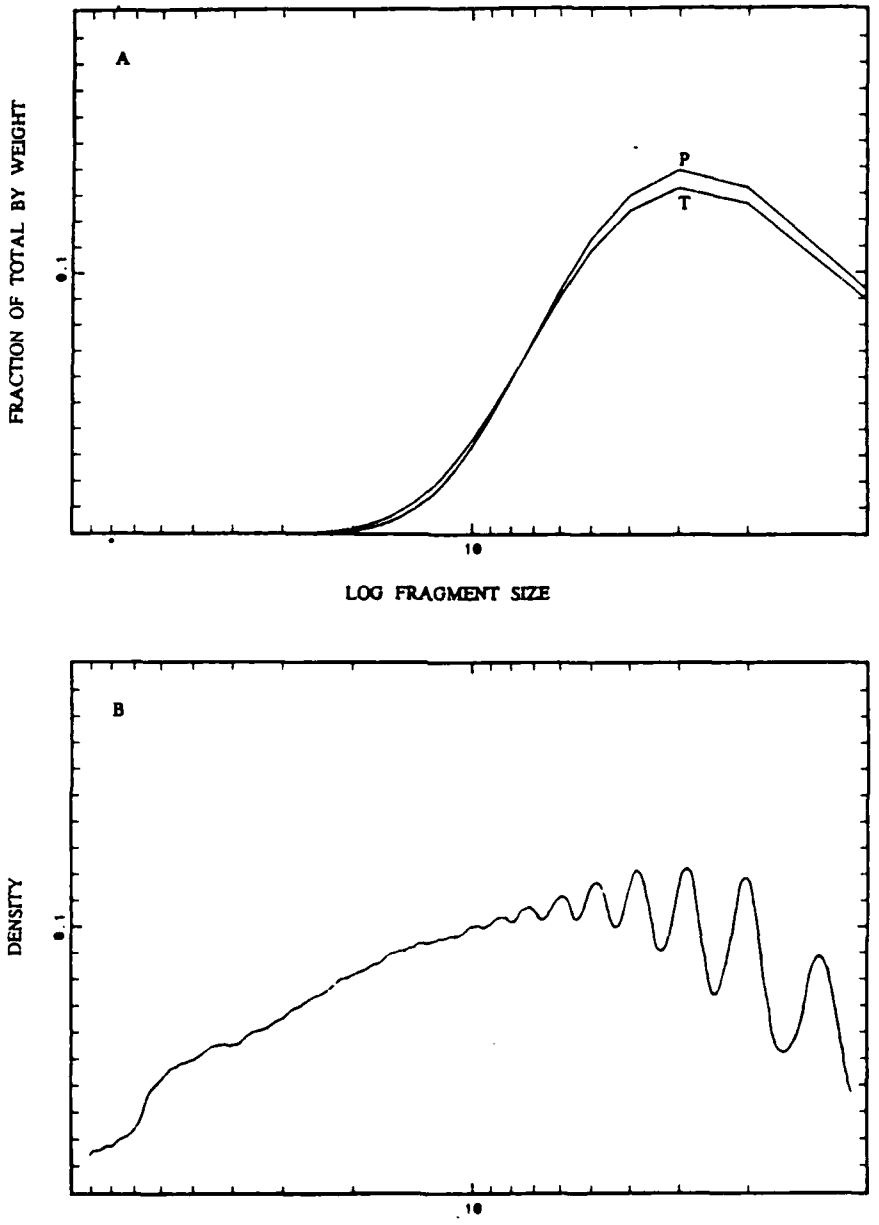


Fig. 3. Model of a Micrococcal nuclease digestion of human chromatin. Panel A: the predicted fragment distribution approximating the Micrococcal nuclease digestion of human bone marrow chromatin shown in panel B. The model assumes a linear substrate of 100 nucleosomes (20 kb) containing equally susceptible nuclease cleavage sites spaced every 200 bp. Panel B: the distribution of digestion products of a Micrococcal nuclease digestion of human bone marrow chromatin described in the materials and methods section. Note that, again, the x-axis is the inverted log of fragment size, in this case, in terms of unit fragment size (200 bp). Thus, the ordinate value of 1 (at the right) represents a mononucleosome of 200 bp.

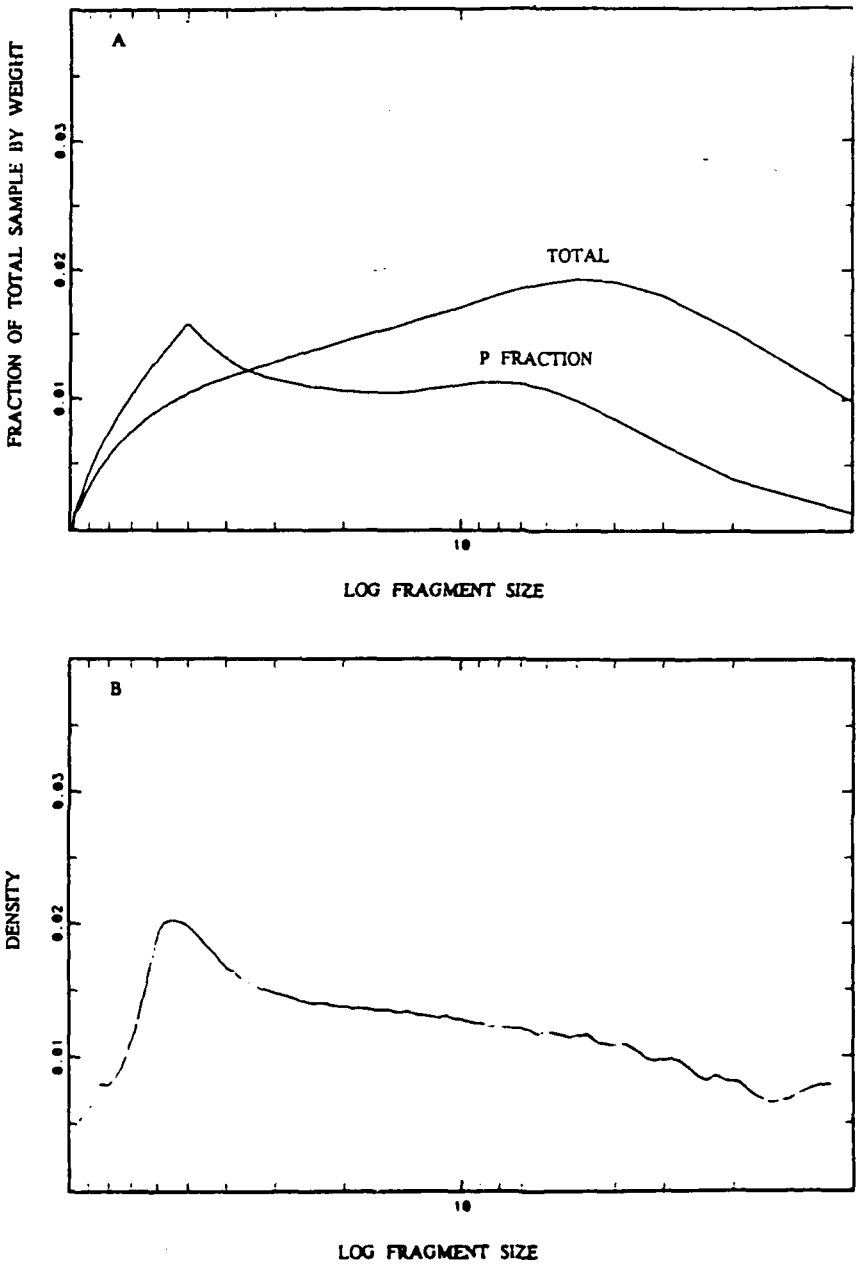


Fig. 4. Model of a fraction of Micrococcal nuclease digested chromatin. Panel A: the predicted fragment distribution of the total and fractionated digestion products obtained from the digestion described in Fig. 3 and the text. The model is identical to that described in Fig 3. except that the distributions of several digests were combined as described in the text. Fractionation of products was modeled by assuming that the center nucleosome was in some way attached to the nuclear matrix. The pellet fraction represents the predicted distribution of products containing the center nucleosome. Panel B: the fragment distribution obtained from a simple fractionation of the chromatin digestion products as described in Fig. 3 and the text. The ordinate is arranged as described in Fig. 3 to simulate the kind of pattern observed in agarose gel electrophoresis.

100 nucleosome model, the predicted distributions of more than one degree of digestion were combined. The distribution obtained using four separate subset digestions of 2, 5, 12, and 30 cleavages weighted 8 : 3 : 2 : 1, respectively, is shown in Fig. 4(A). Without attempting to optimize the fit, this predicted distribution of fragments represented a somewhat better fit of the data than that of Fig. 3.

When isolating chromatin from nuclei rather than simply DNA, a fractionation of the digestion products results from the experimental protocols normally used.[6] In one of these protocols, the human bone marrow nuclei described for Fig. 3 may be removed from the digestion buffer following digestion with Micrococcal nuclease and treated with a very low ionic strength buffer containing EDTA. This causes further release of chromatin fragments from the nuclei. A certain fraction of the chromatin remains in the nuclei, however, depending on the degree of digestion. Further release of chromatin resists additional washing by resuspension and centrifugation. The DNA remaining entrapped in the nuclei shows a complete distribution of fragment sizes, suggesting that size alone is not responsible for its retention. This DNA, called P fraction (Fig. 4), shows that P fraction has a distribution of fragment sizes that is biased in comparison to that of the total digestion products (Fig. 3) toward the high molecular weight range.

In the model calculations performed for the total digest in Fig. 4, a nucleosome in the center of this substrate was earmarked and the distribution of all fragments containing this nucleosome was simultaneously determined. The distribution of these fragments (Fig. 4) also shows good agreement with the experimental results of Fig. 4. That such reasonable predictions may be made with the simple distribution of cleavage states used here demonstrates the utility of modeling complex digestions in chromatin studies. It is interesting to note that these results suggest that the majority of the chromatin has been digested more lightly than the fraction of mononucleosomes would imply. It is also useful to observe that the greatest degree of digestion dominates the distribution of fragments, despite the fact that, in the model digestion, this represents only one-fourteenth of the total chromatin.

Such predictions can also be made using the Tanford model [Eq. (9)]. These are similar to those in Fig. 4 (not shown). A rigorous comparison between models would require experimental data of greater precision. In any case, it is meaningless to compare the models here, since reasonable fits of the experimental data may be obtained for either model simply by adjusting parameters.

DISCUSSION

The use of a 100-nucleosome model to approximate the digestion of a human genome leaves much to be desired. It is justified as a first approximation, however, since at large N the weight fraction of any given fragment size decreases slowly with increasing N for a given degree of digestion. This can be seen from Eqs. (3) and (4). In Tanford's model, weight fraction is a function solely of the degree of digestion, q and fragment length L [(Eq. (9)]. The useful range of resolution of DNA nucleosomal fragments on a normal agarose gel electrophoresis in fact rarely exceeds on the order of 100–200 nucleosomes (20–40 kb). This and the deviation from linearity of migration vs log molecu-

lar weight in the high molecular weight range make comparison of predicted values with experimental data in the high molecular weight range difficult.

End effects must be kept in mind when using small model substrates. The effect of ends in linear substrates is to increase the likelihood that fragments containing them are produced, since only one cleavage rather than two is required. A distribution of fragment sizes for fragments containing an earmarked locus within the substrate, for example (see Fig. 3), will show a discontinuity at the precise fragment size that is unable to contain both the locus of interest and an end. There is no question that this occurs in reality, and may in fact be used to advantage in situations where proximal ends are artificially created in the experiment by a restriction digest, for example. Using small substrate constructs to approximate larger real ones does, however, introduce these end effects where, in reality, they are difficult to detect because of the large fragment molecular weight at which they occur.

Comparisons of the calculated fragment distributions of the Tanford model and the present model show that the differences between the two, for large substrates, are small. At small N, however, where site susceptibility is also important, the two models diverge. The ability of the present model to better approximate a restriction enzyme digestion such as that shown above was demonstrated. In this case, with the treatment of site susceptibility as a pure likelihood rather than a target of some specified size and the concept that each substrate molecule is digested as a discrete entity, the present model may be more useful in describing restriction enzyme digestions than the Tanford model.

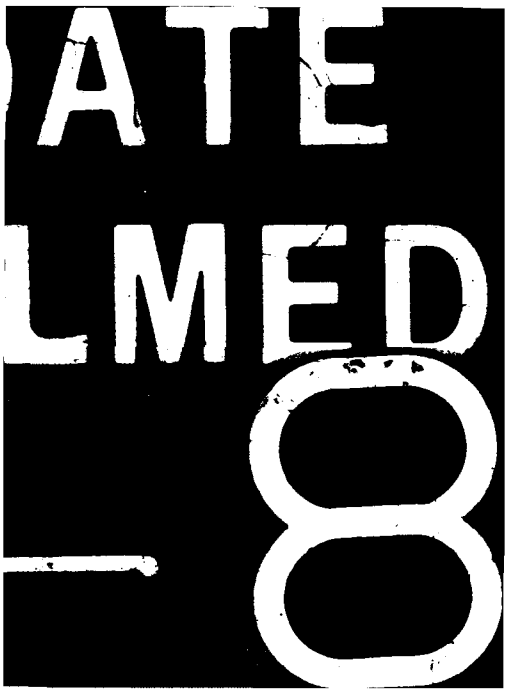
The usefulness of modeling in chromatin studies has not been widely appreciated. There are, nevertheless, many situations in such studies where modeling may be used to advantage for distinguishing between possible explanations of an experimental result or for guidance in experimental design. The interplay of site susceptibility, gene sequence, loci of associated proteins, and nucleosome phasing can be complex, and the use of a model incorporating all of these variables is of great value. This report offers such a model.

This work was supported by the Naval Medical Research and Development Command, Research Task No. MR041.20.03.002. The opinions and assertions contained herein are the solely those of the author and are not to be construed as official or reflecting the views of the Navy Department or the naval service at large.

References

1. Flory, P. J. (1953) *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, pp. 318–331.
2. Tanford, C. (1961) *Physical Chemistry of Macromolecules*, John Wiley & Sons, New York, NY, pp. 138–144, 612–621.
3. Jack, W. E., Terry, B. J. & Modrich, P. (1982) *Proc. Natl. Acad. Sci. USA* 79, 4010–4014.
4. Berg, O., Winter, R. B. & Von Hippel, P. H. (1981) *Biochemistry* 20, 6929–6948.
5. Wallace, P. G., Hewish, D. R., Venning, M. M. & Burgoyne, L. A. (1971) *Biochem. J.* 125, 47–54.
6. Davis, A. H., Reudelhuber, T. L. & Garrard, W. T. (1983) *J. Mol. Biol.* 167, 133–155.
7. Hough, C. J. & Smulson, M. E. (1984) *Biochemistry* 23, 5016–5023.
8. Wood, W. I. & Felsenfeld, G. (1982) *J. Biol. Chem.* 257, 7730–7736.
9. McGhee, J. D., Wood, W. I., Dolan, M., Engel, J. D. & Felsenfeld, G. (1981) *Cell* 27, 45–55.

Received September 16, 1986
Accepted May 25, 1987


ATE
LMED
8